5. Понятие статистической гипотезы и параметрической статистической гипотезы. Понятие критерия проверки гипотез и его задание с использованием критического множества. Описать построение критериев проверки гипотез (а) $H_0 = \{m = m_0\}$, $H_1 = \{m > m_0\}$; (б) $H_0 = \{m=m_0\}$, $H_1 = \{m<m_0\}$; (в) $H_0 = \{m=m_0\}$, $H_1 = \{m\neq m_0\}$ относительно значения m математического ожидания нормальной случайной величины как в случае известной, так и в случае неизвестной дисперсии.

~~~~~~

let X - СВ, з-н распр неизв/изв неполностью с точн до в-ра $\vec{\Theta}$ неизв параметров

- Статической гипотезой назыв любое утв о з распр X
- Стат гипотеза назыв параметрич, если она явл утв отн зн-я $\vec{\Theta}$
- Правило с исп кот прин реш-я об истинности гипотез назыв критерием проверки гипотез. Критерий обычно задается с исп критического мн-ва. $W \subseteq X_n$ (выб пр-во). При этом соотв пр имеет вид: $\vec{x_n} \in W \Rightarrow \{$**отв** H_0**, прин** H_1 $\vec{x_n} \in X \setminus W \Rightarrow \{$**пр** H_0**, отв** H_1. W назаыв доверительным мн-вом. Критерий проверки гип полностью определяется крит мн-вом.
- let X~N(m, σ^2), где m и σ-неизв.
  (а) Рассм з-чу проверки гип $H_0 = \{m = m_0\}$ против в $H_1 = \{m > m_0\}$. В этом примере целесообр воспользоваться статистикой $T(\vec{X}) = \frac{\overline{X} - m_0}{\rho}\sqrt{n}$~ St(n-1). Критик мн-во можно задать в виде
  $W = \{\vec{x} \in \chi_n : T(\vec{x}) \geq t_{1-\alpha}^{(n-1)}\}$, где $t_{1-\alpha}^{(n-1)}$ -квантиль
  (б) $H_0 = \{m=m_0\}$, $H_1 = \{m<m_0\}$.
  (в) $H_0 = \{m=m_0\}$, $H_1 = \{m\neq m_0\}$.
  Рассуждая аналогично, с исп ст-ки $T(\vec{X}) = \frac{\overline{X} - m_0}{\rho}\sqrt{n}$ зададим кричит мн-ва в виде (б)$W = \{\vec{x} \in \chi_n : T(\vec{x}) \leq - t_{1-\alpha}^{(n-1)}\}$,
  (в)$W = \{\vec{x} \in \chi_n : t_{1-\frac{\alpha}{2}}^{(n-1)}\}$
- let X~N(m, σ^2), где m-неизв, σ-изв. (а) $H_0 = \{m=m_0\}$, $H_1 = \{m>m_0\}$.
  $W = \{\vec{x} \in \chi_n : T(\vec{x}) \geq u_{1-\alpha}\}$, где u_{1-a}-квантиль ст н распр $T(\vec{X}) = \frac{\overline{X} - m_0}{\sigma}\sqrt{n}$
  (б) $H_1 = \{m>m_0\}$.
  $W = \{\vec{x} \in \chi_n : T(\vec{x}) \leq - u_{1-\alpha}\}$
  (в) $H_1 = \{m\neq m_0\}$. При истинности конкретн гип-зы H_1 ст-ка T будет принимать больше по абс величине знач-я, поэт критич мн-во можн завис в виде $W = \{\vec{x} \in \chi_n : |T(\vec{x})| \geq u_{1-\frac{\alpha}{2}}\}$

3. Понятие статистической гипотезы и параметрической статистической гипотезы. С использованием критерия Неймана-Пирсона построить критерий проверки двух простых гипотез $H_0 = \{m = m_0\}$, $H_1 = \{m = m_1\}$, $m_1 > m_0$, относительно значения m математического ожидания нормальной случайной величины при известной дисперсии

~~~~~~

let X - СВ, з-н распр неизв/изв неполностью с точн до в-ра $\vec{\Theta}$ неизв параметров

- Статической гипотезой назыв любое утв о з распр X
- Стат гипотеза назыв параметрич, если она явл утв отн зн-я $\vec{\Theta}$

Пусть X~$N(m, \sigma^2)$, где m - неизв, а σ^2 - известно. Рассмотрим задачу проверки двух гипотез $H_0 = \{m = m_0\}$ и $H_1 = \{m = m_1\}$

Функ правдоподобия

$$L(X_1, \ldots, X_n, m) = [\frac{1}{\sqrt{2\pi\sigma}}]^n \cdot e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - m)^2}$$

Тогда отношения правдоподобия:

$$\phi(\vec{X}) = \frac{L(\vec{X}, m_1)}{L(\vec{X}, m_2)} = \frac{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - m_1)^2}}{e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n (x_i - m_0)^2}} =$$

$$= e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [(x_i - m_1)^2 - (x_i - m_0)^2]} =$$

$$= e^{-\frac{1}{2\sigma^2}\sum_{i=1}^n [x_i^2 - 2x_i m_1 + m_1^2 - x_i^2 + 2x_i m_0 - m_0^2]} =$$

$$= e^{\frac{m_1 - m_0}{2\sigma^2}\sum_{i=1}^n x_i - \frac{n}{2\sigma^2}\sum_{i=1}^n (m_1^2 - m_0^2)}$$

Условие $\phi(\vec{X}) \geq c_\phi \Leftrightarrow ln\,\phi(\vec{X}) \geq ln\,c_\phi \Leftrightarrow$

$$\Leftrightarrow ln[e^{\frac{m_1 - m_0}{2\sigma^2}\sum_{i=1}^n x_i - \frac{n}{2\sigma^2}\sum_{i=1}^n (m_1^2 - m_0^2)}] \geq ln\,c_\phi \Leftrightarrow |m_1 > m_0, \text{ поэтому } m_1 - m_0 > 0|$$

$$\Leftrightarrow \sum_{i=1}^n x_i \geq \frac{\sigma^2}{m_1 - m_0}[ln\,c_\phi - \frac{n}{2\sigma^2}(m_1 - m_0)] = c$$

$c = const$. Т.о. $W = \{\vec{x} \in \chi_n : \sum_{i=1}^n x_i \geq c\}$,

где с выбирается из условия

$$\alpha = \{P\{\phi(\vec{x}) \geq c_\phi | H_0\}\} = P\{\sum_{i=1}^n x_i \geq c \,|\, m = m_0\}$$

Если истина H_0, т.е. m=m_0, то сл. вел. $\sum_{i=1}^n X_i \sim N(n\,m_0, n\,\sigma^2)$

$$\text{Т.о. } \alpha = P\{\sum_{i=1}^n x_i \geq c \,|\, m = m_0\} =$$

$$= 1 - P\{\sum_{i=1}^n x_i < c \,|\, m = m_0\} = 1 - \Phi(\frac{c - n\,m_0}{\sqrt{n\,\sigma^2}})$$

т.е. $\Phi(\frac{c - n\,m_0}{\sqrt{n\,\sigma^2}}) = 1 - \alpha$. Т.о.

$\frac{c - n\,m_0}{\sqrt{n\,\sigma^2}} = u_{1-\alpha}$ (квантиль уровня $1 - \alpha$)

$c = \sigma\,u_{1-\alpha}\sqrt{n} - n\,m_0$

Т.о. критерий имеет вид:

$\sum_{i=1}^n x_i \geq \sigma\,u_{1-\alpha}\sqrt{n} + n\,m_0 \Rightarrow$ принять H_1, отклонить H_0

$\sum_{i=1}^n x_i < \sigma\,u_{1-\alpha}\sqrt{n} + n\,m_0 \Rightarrow$ принять H_0, отклонить H_1

1. Понятие статистической гипотезы. Постановка задачи проверки статистических гипотез. Понятие критерия проверки гипотез. Ошибки первого и второго рода, вероятность их совершения. Определение уровня значимости и мощности критерия. Общие принципы построения статистических критериев.

~~~~~~

-Пусть X - СВ, закон распределения которой известен. Статистической гипотезой называют любое утверждение относительно закона распределения СВ X.
- Проверку статистической гипотезы обычно проверяют следующим образом:
1) Формулируют основную гипотезу H_0 2) формулируют конкурирующую(альтернативную) гипотезу H_1. $H_0 \cap H_1 = 0$, но возможно $H_0 \cup H_1$ не исчерпывают всевозможные случаи 3)на основании имеющейся выборки $\vec{X} \in X_n$ принимают решение об истинности H_0 или H_1.
- Правило, посредством которого принимается решение об истинности H_0 или H_1 называют статистическим критерием проверки гипотезы.
- Принимать конкурирующую гипотезу при истинности основной гипотезы — ошибка 1 рода.
$P\{\vec{x} \in W \,|\, H_0\} = \alpha$
принимать H_0 при истинности H_1 — ошибка 2 рода.
$P\{\vec{x} \notin W \,|\, H_1\} = \beta$
- При этом α называют уровнем значимости критерия, а $1 - \beta$ называют мощностью критерия.
- конечно, при построении критерия хотелось бы сделать так, чтобы $\alpha \to \min$, $\beta \to \min$, однако это невозможно, поэтому исходят из принципа
$\beta \to min, \alpha = const$

4. Понятие статистической гипотезы и параметрической статистической гипотезы. Простая и сложная гипотезы. Ошибки первого и второго рода. Вероятности их совершения как функции неизвестного параметра при проверке двух сложных гипотез. Понятия размера критерия и функции мощности. Выражение вероятностей ошибок первого и второго рода через функцию мощности. Понятие равномерно наиболее мощного критерия. Равномерно наиболее мощный критерий при проверке гипотез $H_0 = \{m = m_0\}$, $H_1 = \{m > m_0\}$ относительно значения m математического ожидания нормальной случайной величины при известной дисперсии.

~~~~~~

let X - СВ, з-н распр неизв/изв неполностью с точн до в-ра $\vec{\Theta}$ неизв параметров

- Статической гипотезой назыв любое утв о з распр X
- Стат гипотеза назыв параметрич, если она явл утв отн зн-я $\vec{\Theta}$
- Стат гипотеза назыв простой, если она однозначно определяет закон сл вел X. В противном случае называется сложной.
- Ошибка 1 рода: принять конкурирующую гипотезу при истинности основной. Наоборот - ошибка 2 рода. При этом вероятности их совершения являются функциями от неизвестного параметра:
- $\alpha(\theta) = P\{\vec{X} \in W \,|\, \theta \in (H)_0\}$
- $\beta(\theta) = P\{\vec{X} \in \chi_n \setminus W \,|\, \theta \in (H)_1\}$

При этом θ - неизв пар-р з-на распр , общий вид которого известен с точностью до θ. Где $(H)_0$ - должна быть H в кружочке с нуликом, но тут не такого символа :(

- Величина $a = sup\,\alpha(\theta), \theta \in (H)_0$ наз размером критерия.
- Ф-ция $M(\theta) = P\{\vec{X} \in W \,|\, \theta\}$ нас функцией мощности критерия.
- Выражения вероятн ошибок через функцию мощности:
- $\alpha(\theta) = M(\theta), \theta \in (H)_0$
- $\beta(\theta) = 1 - M(\theta), \theta \in (H)_1$
- Критерий который максим при заданном размере α функцию мощности одновр по всем возможным критериям при всех $\theta \in (H)_1$ наз равномерно наиболее мощным.
  let X~$N(m, \sigma^2)$, где m - неизв, σ^2 -изв, расы задачу проверки гипотез $H_0 = \{m = m_0\}$ - простая, $H_1 = \{m > m_0\}$ - сложная. Для гипотез $H_0 = \{m = m_0\}$ и $H_1 = \{m = m_1\}$, где $m_1 > m_0$ крит мн-во имеет вид: (*)
  $$W = \{\vec{x} \in \chi_n : \sum_{i=1}^n x_i \geq n\,m_0 + u_{1-\alpha}\sigma\sqrt{n}\}$$

Т.к. построенное выше крит мн-во не зав от m_1, то фактически этот критерий явл равн наиб мощным для проверки гипотез $H_0 = \{m = m_0\}$ и $H_1 = \{m > m_0\}$. Т.о. для расм з-чи крит мн-во имеет вид (*)

2. Понятие статистической гипотезы и параметрической статистической гипотезы. Простая и сложная гипотезы. Построить критерий Неймана-Пирсона для проверки двух простых гипотез.

~~~~~~

- Пусть X - СВ, закон распределения которой известен. Статистической гипотезой называют любое утверждение относительно закона распределения СВ X.
- Статистическую гипотезу называют параметрической, если она является утверждением относительно независимых параметров известного закона распределения.
- Статистическую гипотезу называют простой, если она однозначно определяет закон распределения случайной величины X (однозначно задает функцию распределения СВ X как функцию своего аргумента). В протеином случае гипотеза называется сложной.
- Критерий Неймана-Пирсона проверки двух простых гипотез
Пусть:
1) X - СВ 2) $F(x, \theta)$ — функция распределения СВ X (известен общий вид функции F, но она зависит от неизвестного параметра θ)
Рассмотри 2 простые параметрические гипотезы $H_0 = \{\theta = \theta_0\}$ и $H_1 = \{\theta = \theta_1\}$, где $\theta_0 \neq \theta_1$.
Введем на рассмотрение статистику
$\varphi(\vec{x}) = \frac{L(\vec{x}, \theta_1)}{L(\vec{x}, \theta_0})$, где $L(\vec{x}, \theta)$- функция правдоподобия.
Очевидно, что “большие” значения статистики φ ассоциируются с истинностью конкурирующей гипотезы H_1, потому критическое множество должно иметь вид $w = \{\vec{x} \in \chi_n : \varphi(\vec{x}) \geq C_\varphi\}$, где константа C_φ выбирается из условия
$\alpha = P\{\varphi(\vec{x}) \geq C_\varphi | \theta_0 = \theta\}$

———

При построение критерия зафиксировано некоторое $alpha = const$ — вероятность совершения ошибки й рода. Чтобы построенный критерия имел уровень значимости α, необходимо, чтобы $P\{\vec{x} \in W \,|\, H_0\} = \alpha$ в общем случае. В рассматриваемом случае
$\vec{x} \in W \Longleftrightarrow \varphi(\vec{x}) \geq C_\varphi, H_0$ - истина $\Longleftrightarrow \theta = \theta_0$, поэтому
$P\{\vec{x} \in W \,|\, H_0\} = P\{\varphi(\vec{x}) \geq C_\varphi | \theta_0 = \theta\}$

———

Замечание:
1) Построенный критерий называют критерием Неймана-Пирсона
2) Если X - непрерывная СВ, то плотность распределения сл выборки $\vec{X}$ имеет вид:
$f_{\vec{x}}(x_1, \ldots, x_n, \theta) = f(x_1, \theta) \cdot \ldots \cdot f(x_n, \theta) = L(x_1, \ldots, x_n, \theta)$
,где $f(x, \theta)$ — функция плотности распределения СВ X.

7.Постановки первой и второй основных задач математической статистики. Основные идеи решения первой задачи. Определение критерия согласия. Основные принципы, используемыепри формулировке основной гипотезы при решении первой задачи.

~~~~~~

• Вторая задача МС: Дано СВ X, функция $F(x, \vec{\theta})$ распределния, которой известна с точностью до вектора $\vec{\theta}=(\theta1,...,\theta n)$ неизвестных параметров. Требуется оценить значение вектора $\vec{\theta}$.
• Первая задача МС. Дано СВ X, закон распределения которой неизвестен. Требуется определить функцию распределения СВ X.
Решение этой задачи сводится к проверке основной гипотезы $H0=\{F(t)\equiv F0(t, \vec{\theta})\}$($F$ - функция распределения СВ X, $F0$ - нек. известная функция распределения) против конкурирующей гипотезы $H1=\neg H0=\{(\exists t\in R)(F(t)\neq F0(t))\}$ Проверка основной гипотезы $H0$ сводится к оценке величины $\Delta(Fn,F0)$ рассогласована эмпирической функции распределения Fn предполагаемой функции распределения $F0(t, \vec{\theta})$.
• Критерием согласия называют статистические критерии, предназначенные для проверки корректности гипотезы о том что предполагаемый закон распределения $F0(t, \vec{\theta})$ соответствует экспериментальным данным представленным эмпирической функцией распределения $Fn(t)$.
• При выдвижении основной гипотезы $H0$ может быть истинны следующие соображения:
1) Анализ априорной информации об изученном явлении и ее сопоставлении с исходными предпосылками построения конкретных моделей (норм, экспоненц)
2) Построение эмпирической функции распределения по данным выборки $\vec{X}$n. На основе вида этой функции может быть выдвинута основная гипотеза.
3) Построение гистограммы по данным выборки $\vec{X}$ n. На ее основании может быть выдвинута гипотеза о виде закона распределения X.

8.Постановка задачи проверки гипотезы о законе распределения случайной величины. Описать критерий Колмогорова для проверки простой гипотезы. Сформулировать утверждения о законе распределения соответствующей статистики.

~~~~~~

• Задачу проверки стат. гипотез обычно ставят следующим образом:
- проверяемую гипотезу H_0 наз. осн. или нулевой
- формулируют гип-зу H_1, которую наз. альтернативной или конкурирующей.
При этом эти гипотезы не должны пересекаться $H_0 \cap H_1$ =0, но возможн H_0 и H_1 не исчерпывают все возможн случаи.
• Критерий Колмогорова для простой гипотезы
Пусть x – сл вел, $\vec{x}$ - случайная выборка из ген. совокупности X
Рассм задачу проверим гипотезы $H_0=\{F(t)= F_0(t)\}$ против $H_1 = \neg H_0$
Где F0 (t) – полностью известная ф-ия распр, H_0– простая гипотеза
Рассм. ст-ку Δ, реализ. котор. опр. соотношением
Очевидно, что “малые” знач. ст-ки $\Delta(\vec{x})$ свидетельств об ист-ти H_0,
а “большие” об ист-ти H_1
По этой причине критич мн-во имеет вид $W=\{\vec{x}\in \chi_n: \Delta(\vec{x}) \geq \sigma_{1-\alpha}\}$
Где $\sigma_{1-\alpha}$-квантиль уровня $1 - \alpha$ з-на распр. сл. вел $\Delta(\vec{x})$
При этом собств решающее правило им. сл. вид:
$\vec{x}\in W \Rightarrow$ прин H_1, откл H_0
$\vec{x}\in W \Rightarrow$ прин H_0, откл H_1
• Замеч о законе распр сл вел $\Delta(\vec{x})$
1) Построенный критерий предпологает исп-е квантилей зак распр сл. вел. $\Delta(\vec{x})$
2) В ст-ке доказано утв-е
Пусть: a) Y~R[0;1]
b) $\hat{R}n(t, \vec{Y})$– выборочная ф-ия распр t сл. вел. Y
Тогда при ист-ти H_0 ф-ия распр сл. вел. $\Delta(\vec{x})$ совпадает с ф-ей распр сл. вел.
$Z(u)=\sup|\hat{R}n(t, \vec{Y})- t|$; $t\in[0;1]$
3) Для каждого n ∈ N з-н распр сл.в. Z(n) известен
Для n ≤ 100 составлены табл знач соотв ф-ий распр
4) Колмогоров доказал что
$$\lim_{n\to\infty} P\left\{\sqrt{n}\,Z(n) < t\right\} = K(t);\ t > 0$$
где $K(t)= \sum_{n=-\infty}^{+\infty} (-1)^k e^{-2k^2t^2}$
5) Как показывает практика, соотн-м $\sigma_{1-\alpha}=\frac{a_{1-\alpha}}{\sqrt{n}}$ можно польз при n ≥ 20
~~~~~~

13. Основные задачи регрессионного анализа. Понятие модели, линейной по параметрам. Определение МНК-оценки параметров и ее вычисление (без доква). Теорема о свойствах построенной оценки.

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

Пусть есть результаты n наблюдений в виде:

$$\begin{cases} y_1 = \Phi(x_1) + \xi_1 \\ \dots \\ y_n = \Phi(x_n) + \xi_n \end{cases}, \text{ где}$$

1)$y_1, \dots, y_n$ - n реализаций Y;

2)$\xi_1, \dots, \xi_n$ - n релизаций ξ;

3)$x_1, \dots, x_n$ - известные значения.

Требуется на основе этих данных подобрать $\hat\Phi$ так, чтобы она наилучшим образом аппроксимировала неизвестную функцию Φ.

Часто в качестве функции $\hat\Phi(x)$ выбирают функцию следующего вида:

$\hat\Phi(x) = \theta_1\psi_1(x) + \dots + \theta_p\psi_p(x)$, где

$\psi_1(x), \dots, \psi_p(x)$ - известные функции.

Параметры $\theta_1, \dots, \theta_p$ подбирают так, чтобы $\hat\Phi(x)$ наилучшим образом аппроксимировала $\Phi(x)$.

Таким образом результаты наблюдений можно записать в виде:

$$y_i = \theta_1\psi_1(x_i) + \dots + \theta_p\psi_p(x_p) + \xi_i,\ i = \overline{1,n}$$

В матричном виде:

$\vec y = \Psi\vec\theta + \vec\xi$, где

$$\vec y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$$

$$\Psi = \begin{pmatrix} \psi_1(x_1), \cdots, \psi_p(x_1) \\ \vdots\ ,\ \ddots\ ,\ \vdots \\ \psi_1(x_n), \cdots, \psi_p(x_n) \end{pmatrix}\ \vec\theta = \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_n \end{pmatrix},\ \vec\xi = \begin{pmatrix} \xi_1 \\ \vdots \\ \xi_n \end{pmatrix}$$

Задача заключается в подборе $\vec\theta$.

Предполагается, что систематические ошибки отсутствуют, то есть $M\xi = 0, \xi \sim N(0,\sigma^2)$.

Оценка $\hat{\vec\theta}$ вектора $\vec\theta$ называется оценкой, полученной методом наименьших квадратов (МНК-оценкой), если $S(\vec\theta) = \|\vec y - \Psi\vec\theta\|^2$ - минимально.

В рассматриваемом случае МНК-оценка вектора $\vec\theta$ имеет вид:

$\hat{\vec\theta} = (\Psi^T\Psi)^{-1}\cdot\Psi^T\vec y$, где $rg(\Psi) = p$.

Причем так как $y = \theta_0 + \theta_1 t + \theta_2 t^2$, то

$$\Psi = \begin{pmatrix} 1 & t_1 & t_1^2 \\ 1 & t_2 & t_2^2 \\ \vdots & \vdots & \vdots \\ 1 & t_1 & t_1^2 \end{pmatrix}$$

Среднеквадратичное отклонение полученной модели от результатов наблюдений будем вычислять как $\Delta = \sqrt{\sum_{i=1}^{n}(y_i - y(t_i))^2}$.

9. Постановка задачи проверки гипотезы о законе распределения случайной величины. Описать критерий χ 2 для проверки простой гипотезы. Сформулировать утверждения о законе распределения соответствующей статистики.

Задачу проверки стат. гипотез обычно ставят следующим образом:

- проверяемую гипотезу H_0 наз. осн. или нулевой
- формулируют гип-зу H_1, которую наз. альтернативной или конкурирующей.

При этом эти гипотезы не должны пересекаться $H_0 \cap H_1$ =0, но возможн H_0 и H_1 не исчерпывают все возможн случаи.

Для критерия χ 2 для проверки простой гипотезы задача форм сл образом:

Пусть X – дискр. величина, которая может принимать знач $a_1..a_L$ с вер-ю $p_1..p_L$ соотв. Требуется проверить гипотезу:

$H_0 = \left\{p_i = p_i^0,\ i = \overline{1,L}\right\}$ против

$H_1 = \neg H_0 = \left\{\exists i \in \{1, \dots, L\}\left(p_i \ne p_i^0\right)\right\}$

$$\sum_{i=1}^{L} p_i^0 = 1$$

Для решения этой задачи введем статистику $n_i\left(\vec X\right),\ i = \overline{1:L}$, выборочные значения которых опр след образом:

$n_i\left(\vec x\right) = \left\{ \begin{array}{l} кол-во\ компонент\ вектора\ \vec x, \\ кот\ приняли\ значение\ a_i \end{array} \right\}$

$i = \overline{1,L}$

Теорема Пирсона. Пусть выполнены сделанные выше предположения. Тогда при истинности H_0 посл-ть сл величин

$$\sum_{i=1}^{L}\frac{\left(n_i\left(\vec x\right) - n p_i^0\right)^2}{n p_i^0} \text{ при } n \to \infty \text{ слабо}$$

сходится к сл вел, имеющей распр χ 2(L-1).

Согласно этой теореме при $n \to \infty$ статистика

$$\Delta_n\left(\vec x\right) = \frac{n^2}{n}\sum_{i=}^{L}\frac{\left(\frac{n_i\left(\vec x\right)}{n} - p_i^0\right)^2}{p_i^0}$$

При этом «малые» знач стат Δ_n ассоциируется с истинностью осн гипотезы H_0, а «большие» - с истинностью H_1. По этой причине критич мн-во можно задать в виде

$w = \left\{\vec x_n \in \chi_n: \Delta_n\left(\vec x_n\right) \ge h_{1-\alpha}(L-1)\right\}$, где h – квантиль ур-я $1-\alpha$ распределения χ 2(L-1).

10. Постановка задачи проверки гипотезы о законе распределения случайной величины. Описать критерий Колмогорова для проверки сложной гипотезы. Трудности, связанные с использованием этого критерия и их преодоление.

Задачу проверки стат. гипотез обычно ставят следующим образом:

- проверяемую гипотезу H_0 наз. осн. или нулевой
- формулируют гип-зу H_1, которую наз. альтернативной или конкурирующей.

При этом эти гипотезы не должны пересекаться $H_0 \cap H_1$ =0, но возможн H_0 и H_1 не исчерпывают все возможн случаи.

Для критерия Колмогорова для проверки сложной гипотезы задача форм сл образом:

Пусть X – непр сл вел. На практике задача проверки гипотезы $H_0 = \{F(t) \equiv F_0(t)\}$ против $H_1 = \neg H_0$ т.о. приходим к задаче проверки гипотезы

$H_0 = \left\{\left(\exists\vec\theta\right)\left(F(t) \equiv F_0\left(t, \vec\theta\right)\right)\right\}$ против гип

$H_1 = \neg H_0 = \left\{\left(\forall\vec\theta\right)(\exists t)\left(F(t) \ne F_0\left(t, \vec\theta\right)\right)\right\}$

Для решения этой задачи кажется естественным сделать следующее:

1.Построить точечную оценку $\vec\theta\left(\vec x\right)$ значения вектора параметров $\vec\theta$

2.Использовать критерий Колмогорова для проверки простой гипотезы.

$H_0 = \left\{F(t) \equiv F_0(t),\ \vec\theta\left(\vec x\right)\right\}$, где $\vec\theta\left(\vec x\right)$ – выборочное значение статистики $\hat\theta\left(\vec x\right)$

Недостатком этого подхода явл то, что в этом случае критерий перестает быть параметрическим, т.к. распр модиф статистики

$\hat\Delta\left(\vec x\right) = \sup_{t \in R}\left|F(t) - F_0\left(t, \hat\theta\left(\vec x\right)\right)\right|$ при ист-ти H_0 зависим, вообще говоря, не только от выбора ф-ии F_0, но и от способа построения точечной оценки $\vec\theta$.

Однако, можно показать, что если

1)$\hat\theta\left(\vec x\right)$ – оценка max правдоподобия

2)Элементы $F_0\left(t, \vec\theta\right)$ параметрического семейства получаются с исп-м сдвига и масштаба какого-нибудь одного своего представления, т.е.

$$F_0\left(t, \vec\theta\right) = F_0\left(\frac{t-a}{b}, \vec\theta\right)$$

То для исп-я крит Колмогорова достаточно иметь лишь одну таблицу квантилей для каждого семейства (их распределения). Также табл составл для часто встречающихся семейств (нормального, экспоненциального)

6. Понятие статистической гипотезы и параметрической статистической гипотезы. Понятие критерия проверки гипотез и его задание с использованием критического множества. Опи- сать построение критериев проверки гипотез (а) $H_0 = \{m_1 = m_2\}$, $H_1 = \{m_1 > m_2\}$; (б) $H_0 = \{m_1 = m_2\}$, $H_1 = \{m_1 < m_2\}$; (в) $H_0 = \{m_1 = m_2\}$, $H_1 = \{m_1 /= m_2\}$ относительно значений m_1 и m_2 математических ожиданий двух независимых нормальных случайных величин как в случае известных, так и в случае неизвестных дисперсий.

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

let X - СВ, з-н распр неизв/изв неполностью с точн до в-ра $\vec\Theta$ неизв параметров

- Статической гипотезой назыв любое утв о з распр X
- Стат гипотеза назыв параметрич, если она явл утв отн зн-я $\vec\Theta$
- Правило с исп кот прин реш-я об истинности гипотез назыв критерием проверки гипотез. Критерий обычно задается с исп критического мн-ва. $W \subseteq X_n$ (выб пр-во). При этом соотв пр имеет вид: $\overline{x_n} \in W \Rightarrow \{$**отв** H_0**, прин** H_1 $\overline{x_n} \in X \setminus W \Rightarrow \{$**пр** H_0**, отв** H_1. W назаыв доверительным мн-вом. Критерий проверки гип полностью определяется крит мн-вом.
- let 1) X~N(m_1, σ_1^2); 2) Y~N(m_2, σ_2^2) ;3) m1,m2-неизв, σ1,σ2-изв. Рассм з-чу проверки гип:
  (а) $H_0 = \{m_1 = m_2\}$ против $H_1 = \{m_1 > m_2\}$
  (б)$H_0 = \{m_1 = m_2\}$, $H_1 = \{m_1 < m_2\}$
  (в) $H_0 = \{m_1 = m_2\}$, $H_1 = \{m_1 \ne m_2\}$; Рассм сл вел Z=X-Y; MZ=MX-MY поэтому сформулируем з-чу эквив з-м: (а) H_0 = {m=0}, H_1 = {m>0} (б) H_0 = {m=0}, H_1 = {m<0} (в) H_0 = {m=0}, H_1 = {m≠0}, где m=M[Z]. Рассм стат-ку
  $T(\vec X, \vec Y) = \frac{\overline X - \overline Y}{\sqrt{\frac{\sigma_1^2}{n_1} - \frac{\sigma_2^2}{n_2}}}$, где n1-объев выборки $\vec X$, n2- -//- $\vec Y$. T явл ЛК норм сл вел=>T сама имеет норм распр. M[T]=$\frac{1}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$
  *(M[$\overline X$] - M[$\overline Y$])=|при ист-ти H0 m1=m2|=0
  $D[T] = \frac{1}{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} * [D\overline X + D\overline Y] =$
  $= \frac{1}{\frac{\sigma_1^2}{n_1} + \sigma_2^2} * [\frac{\sigma_1}{n_1} + \frac{\sigma_2}{n_2}] = 1$. Т.о. при ист-ти H0 ст-ка T~N(0,1). По этой причине критич мн-ва в кажд, из расс-х зн-ч имеют вид:
  (а) W={$\vec x, \vec y \in \chi_n : T(\vec x, \vec y) \ge u_{1-\alpha}$},m>0
  (б) W={$\vec x, \vec y \in \chi_n : T(\vec x, \vec y) \le -u_{1-\alpha}$}, m<0
  (в) W={$\vec x, \vec y \in \chi_n : T(\vec x, \vec y) \ge u_{1-\frac{\alpha}{2}}$},m≠0
- let 1) X~N(m_1, σ_1^2); 2) Y~N(m_2, σ_2^2) ;3) m1,m2-неизв, σ1,σ2-неизв. Рассм з-чу проверки гип:
  (а) $H_0 = \{\sigma_1 = \sigma_2\}$ против $H_1 = \{\sigma_1 > \sigma_2\}$
  (б)$H_0 = \{\sigma_1 = \sigma_2\}$, $H_1 = \{\sigma_1 < \sigma_2\}$
  (в) $H_0 = \{\sigma_1 = \sigma_2\}$, $H_1 = \{\sigma_1 \ne \sigma_2\}$. Рассм стат-ку
  $T(\vec X, \vec Y) = \frac{S^2(\vec X)}{S^2(\vec Y)}$ ~(при истинности H_0) $F(n_1, n_2)$. Рассуждая аналогично предыдущему примеру
  (а) W={$\vec x, \vec y \in \chi_n : T(\vec x, \vec y) \ge f_{1-\alpha}$}
  (б) W={$\vec x, \vec y \in \chi_n : T(\vec x, \vec y) \le f_{\alpha}$}
  (в) W={$\vec x, \vec y \in \chi_n : (T(\vec x, \vec y) \le f_{\frac{\alpha}{2}}) \cup$
  $\cup (T(\vec x, \vec y) \ge f_{1-\frac{\alpha}{2}})$}, где f_q-квантиль ур-я распределения Фишера с числом степеней n_1, n_2

11. Постановка задачи проверки гипотезы о законе распределения случайной величины. Описать критерий χ^2 для проверки сложной гипотезы. Сформулировать утверждения о законе распределения соответствующей статистики. Построение оценки максимального правдоподобия в рассматриваемом случае.

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

• Задачу проверки стат. гипотез обычно ставят следующим образом: - проверяемую гипотезу H_0 наз. осн. или нулевой - формулируют гип-зу H_1, которую наз. альтернативной или конкурирующей. При этом эти гипотезы не должны пересекаться $H_0 \cap H_1$ = 0, но возможн H_0 и H_1 не исчерпывают все возможн случаи.

• Пусть 1) X – дискр сл вел, принимающая зн-я $a_1, \dots, a_l$ с вер-ми $p_1, \dots, p_l$ соответственно (v_i – неизв) 2) Ф-я распр-я св X завсисит от в-ра $\vec\theta$ неизв пар-ров, $\vec\theta \in$ *Н в кружочке* Тогда вер-ти $p_i = p\{X = a_i\}$ зависят от зн-я в-ра $\vec\theta$, т е $p_i = p(\vec\theta)\ i = \overline{1,l}$ Т.к. эти зн-я полностью опр-т з-ы распр-я св X, то основную гип-зу можно записать в виде $H_0 = p_i \equiv p_i^0(\vec\theta),\ i = \overline{1,l}\}$, где $p_i^0(\vec\theta)$ -известная ф-ии. При этом $H_1 = \neg H_0$. Для проверки этой сложной гипотезы H_0 исп-ют модифицированный критерий χ^2. 1) Сначала строят оценку макс. Правдоподобия для в-ра $\vec\theta$ пар-ов: $\hat{\vec\theta} = \hat{\vec\theta}(\vec X)$. Вычисляют выборочные зн-я $\hat\theta\left(\vec X\right)$ и $n_i\left(\vec X\right),\ i = \overline{1,l}$ (n_i – тоже, что и в п. 3). 3) Рассматривают статистику

$$\hat\chi^2\left(\vec x\right) = \sum_{i=1}^{l}\frac{\left[n_i\left(\vec X\right) - n p_i^0(\hat{\vec\theta}(\vec X))\right]^2}{n p_k^0(\hat\theta\left(\vec X\right))} =$$

$$= n\sum_{i=1}^{l}\frac{\left[n_i\left(\vec X\right)/n - p_i^0(\hat{\vec\theta}(\vec X))\right]^2}{p_i(\hat\theta\left(\vec X\right))},$$

к-рая при ист-ти H_0, вып-т некот усл гладкости ф-ий p_i^0 при n → ∞ слабо сх-тся к св, имеющей распр $\chi^2(l-r-1)$, где r – размерность в-ра $\vec\theta$.

4) Поскольку с истинностью осн гип-зы H_0 ассоциируются “малые” зн-я ст-ки $\hat\chi^2$, а с истинностью H_1 большие её зн-я, то критич мн-во можно записать в виде

$W = \{\vec x : \hat\chi\left(\vec x\right) \ge h_{1-\alpha}^{(l-r-1)}\}$, где $h_{1-\alpha}^{(l-r-1)}$ – квантиль ур-я $1-\alpha$ распр-я $\chi^2(l-r-1)$.

• При ист-ти H_0 ф-я правдоподобия имеет вид

L($\vec X, \vec\theta$) = $\prod_{k=1}^{n} p\{X = X_k\}$ = |т.к. среди эл-ов в-ра $\vec X$ зн-е a_k встречается n раз, а эл-ты в-ра $\vec X$ могут быть разл способами раскиданы по зн-м a_k.| =

$\frac{n!}{n_1!*\dots*n_l!}\prod_{k=1}^{l} p_k^0(\vec\theta)^{n_k(\vec X)}$, где $\sum_{i=1}^{l} n_k(\vec X) = n$.

12. Постановка задачи о проверке гипотезы о совпадении законов распределения двух случайных величин. Описать критерий Смирнова для решения этой задачи. Сформулировать утверждения о законе распределения соответствующей статистики.

~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~

• Пусть 1) X, Y – св 2) F(t) – ф-я распр св X G(t) - --//-- Y 3) $\vec X$ – выборка из ген сов X (объема n_l) $\vec Y$ – сл -//- Y (--//--n_2) Треб-ся проверить гип-зу H_0 = {X и Y одинаково распр-ны} = $\{\forall t \in R)(F(t) = G(t))\}$ против $H_1 = \neg H_0$. Если св X и Y непр-ны, то для реш-я этой з-чи можно исп-ть ст-ку $\Delta(\vec X, \vec Y)$ выборочн зн-я которой опр-ся ф-лой

$\Delta(\vec x, \vec y) = \sup|F_{n1}(t) - G_{n2}(t)|,\ t \in R$ где $F_{n1}(t)$ и $G_{n2}(t)$ – эмпирич ф-ии распр-я, отвечающие выборкам $\vec x$ и $\vec y$. Если истина H_0, то в соотв с th о сходимости эмпирич ф-ии распр к теоретич ф-ии распр заключаем, что при достаточно больших $n1$ и $n2$ знач-я статистики должны быть “малыми”, а при ист-ти H_1 – “большими”. По этой причине критич мн-во можно задать в виде: W = $\{(\vec x, \vec y): \Delta(\vec x, \vec y) \ge S_{1-\alpha}\}$, где $\alpha \in (0,1)$ – заданный уровень знач-ти критерия. $S_{1-\alpha}$ – квантиль ур-ня 1- α з-на распр ст-ки Δ при ист-ти H_0

• Построенный критерий наз критерием Смирнова

О з-не распр-я статистики $\Delta(\vec X, \vec Y)$:

а) Доказано, что при истин-ти H_0 з-н распр стат Δ не завсисит от F(t) – теоретич (т.е. “истинного”) з-на распр св X (поскольку H_0 предполагается истинной, X и Y имеют одинак распр) б) Для небольших $n1$ и $n2$ соотв распр табулированы (т.е. составлены таблицы их квантилей) в) Смирнов доказал, что для t > 0

P{ $\sqrt{\frac{n1n2}{n1+n2}}\Delta(\vec X, \vec Y) < t\} \to k(t)$ (n р и $n1 \to \infty$ $n2 \to \infty$), где $k(t) = \sum_{k=-\infty}^{+\infty}(-1)^k e^{-2k^2t^2}$ Другими словами при дост-но больших $n1$ и $n2$ можно считать, что св

$A = \sqrt{\frac{n1n2}{n1+n2}}\Delta(\vec X, \vec Y)$ имеет свой з-н распр k(t), t > 0 (т.к. из опр стат Δ вытекает, что $\Delta \ge 0$, то $A \ge 0$ $\Rightarrow F_A(t) \equiv 0, t \le 0$)